#4
~~0240~~
0300

OIPE JC139
OCT 2 6 2000
PATENT & TRADEMARK OFFICE

169.1826

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	:	Examiner: N.Y.A.
ERNEST YIU CHEONG WAN	)	
	:	Group Art Unit: N.Y.A.
Application No.: 09/662,705	)	
	:	
Filed: September 15, 2000	)	
	:	
For: METHOD AND SYSTEM FOR ADDRESSING AUDIO-VISUAL CONTENT FRAGMENTS	) :)	October 25, 2000

Commissioner for Patents
Washington, D.C. 20231

CLAIM TO PRIORITY

Sir:

Applicant hereby claims priority under the International Convention and all rights to which he is entitled under 35 U.S.C. § 119 based upon the following Australian Priority Application:

PQ3122, filed September 27, 1999.

A certified copy of the priority document is enclosed.

Applicant's undersigned attorney may be reached in our New York office by telephone at (212) 218-2100. All correspondence should continue to be directed to our address given below.

Respectfully submitted,

Attorney for Applicant

Registration No. 29,296

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200

NY_MAIN 121138 v 1

OIPE JC139
OCT 2 6 2000
PATENT & TRADEMARK OFFICE

AUSTRALIA

Patent Office
Canberra

I, LEANNE MYNOTT, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ3122 for a patent by CANON KABUSHIKI KAISHA filed on 27 September 1999.

I further certify that pursuant to the provisions of Section 38(1) of the Patents Act 1990 a complete specification was filed on 19 September 2000 and it is an associated application to Provisional Application No. PQ3122 and has been allocated No. 59475/00.

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

WITNESS my hand this
Fifth day of October 2000

LEANNE MYNOTT
<u>TEAM LEADER EXAMINATION</u>
<u>SUPPORT AND SALES</u>

CERTIFIED COPY OF PRIORITY DOCUMENT

S & F Ref: 480155

ORIGINAL

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION FOR THE INVENTION ENTITLED:

Method and System for Addressing Audio-Visual Content Fragments

---

Name and Address of Applicant: Canon Kabushiki Kaisha, incorporated in Japan, of 30-2, Shimomaruko 3-chome, Ohta-ku, Tokyo, 146, JAPAN

Name of Inventors: Ernest Yiu Cheong Wan and Alison Joan Lennon

This invention is best described in the following statement:

5805

Method and System for Addressing Audio-Visual Content Fragments

Technical Field of the Invention

The present invention relates generally to retrieval of data from data bases, and in particular, to retrieval of audio-visual data.

Background Art

The advent of technology providing mass-market access to the Internet places vast amounts of on-line information within relatively easy reach. The World Wide Web (WWW) (hereunder, the Web) underpins much of the growth of Internet use, particularly because of the ease of use, and also due to the intuitive user interface presented by Web browsers. Universal Resource Indicators (URIs) are a ubiquitous addressing feature used to locate target resources in the Web context. This is particularly relevant when Web pages are used in conjunction with a Common Gateway Interface (CGI) scripting application, which allows the Web page to become, in essence, the front end of a myriad of databases accessible over the Internet.

Notwithstanding the explosive progress described however, a Web user is, in most cases, unable to "drill down" beyond a certain level of data, and must, in many cases, down-load an inconveniently large and cumbersome amount of information in order to locate useful information. Illustrating this fact, consider investigating all flights from London to Moscow departing from Heathrow airport on a given date. In order to make a selection based on a number of criteria such as departure time, airline, number of stops and so on, a long list of flights typically needs to be down-loaded and scanned, either manually or using a back-end application on a local personal computer (PC).

Further exemplifying the problem, certain types of data such as, for example, audio-visual (AV) data, typically manifest themselves as monolithic blocks of information. The internal structure of such data, whether it be a particular video segment,

or fragment, in a movie, or a specific movement in a symphony, is neither visible, nor addressible, or consequently accessible in terms of fragments.

Taking a more extreme example, off-line audio-visual data, in the form of celluloid film archives, paper-based libraries, and a wealth of other sources, are also not addressable, and are thus invisible and inaccessible at the "fragment" level. Although particular books can be located, by call number and location in a library, specific chapters thereof are not visible or addressable, and consequently, not accessible.

Extensible Markup Language (XML) provides a drill down capability for a limited sub-set of on-line information, namely information which is coded in XML. However, useful as this may be, the overwhelming bulk of available information has been produced in other programming formats such as Hypertext Markup Language (HTML), or alternatively, is in hard copy form in physical archives and libraries. The aforementioned types of information are referred to as "legacy" information.

It is an object of the present invention to substantially overcome, or at least ameliorate, one or more disadvantages of existing arrangements.

Disclosure of the Invention

It is an object of the present invention to substantially overcome, or at least ameliorate, one or more disadvantages of existing arrangements.

According to a first aspect of the invention, there is provided a method for addressing an arbitrary fragment of an audio-visual (AV) resource belonging to a class of AV resources, the method comprising the steps of:

identifying a logical model for the class of AV resources;

applying the model to the AV resource to form a hierarchical representation of said AV resource including a representation of the AV fragment;

determining a first address for the AV resource;

determining a second address for the representation of the AV fragment depending upon the hierarchical representation; and

combining the first and second addresses to determine an address for the AV fragment.

According to a second aspect of the invention, there is provided a system for addressing an arbitrary fragment of an audio-visual (AV) resource belonging to a class of AV resources, the system comprising:

identification means for identifying a logical model for the class of AV resources;

application means for applying the model to the AV resource to form a hierarchical representation of said AV resource said AV resource representation including an associated root node and a representation of the AV fragment;

first determination means for determining a first address for the AV resource root node;

second determination means for determining a second address for the representation of the AV fragment depending upon the hierarchical representation; and

combining means for combining the first and second addresses to determine an address for the AV fragment.

According to a third aspect of the invention, there is provided a method for addressing an arbitrary fragment of an audio-visual (AV) resource belonging to a class of AV resources, the method comprising the steps of:

determining a first address for the AV resource; characterised in that the method identifies a logical model for the class of AV resources, whereby applying the model to the AV resource forms a hierarchical representation of said AV resource including a representation of the AV fragment, the method comprising the further steps of;

determining a second address for the representation of the AV fragment depending upon the hierarchical representation; and

combining the first and second addresses to determine an address for the AV fragment.

According to a fourth aspect of the invention, there is provided a method for addressing an arbitrary fragment of an audio-visual (AV) data set, whereby application of a first logical model to the AV data set according to a first set of rules has formed a one-to-one meta-data representation of the AV data set, said representation of the AV data set including at least a meta-data representation of said fragment, the meta-data representation of the fragment being associated with a reference, said method comprising the steps of:

selecting the reference associated with the meta-data representation of the fragment; and

applying a second logical model to the selected reference according to a second set of rules to form a meta-data path pointing to the fragment.

According to a fifth aspect of the invention, there is provided a system for addressing an arbitrary fragment of an audio-visual (AV) data set, said system comprising:

first application means for applying a first logical model to the AV data set according to a first set of rules to form a one-to-one meta-data representation of the AV data set, said representation of the AV data set including at least a meta-data representation of said fragment, said meta-data representation of the fragment being associated with a reference;

selection means for selecting the reference associated with said meta-data representation of the fragment; and

second application means for applying a second logical model to the selected reference according to a second set of rules to form a meta-data path pointing to the fragment.

Brief Description of the Drawings

Various aspects of the prior art, and a preferred embodiment of the present invention will now be described with reference to the drawings, in which:

Fig. 1 depicts a prior art system for accessing audio data on a CD ROM using the Internet;

Fig. 2 illustrates indexing typically provided for a CD ROM according to Fig. 1;

Fig. 3 depicts a preferred embodiment of the addressing method in relation to CD ROMs according to the present invention;

Fig. 4 illustrates application of the method in Fig. 3 to addressing a fragment of audio data on a CD ROM;

Fig. 5 depicts the preferred embodiment applied to addressing a fragment of digital video content on a CD ROM;

Fig. 6 depicts the locating of resources using conventional URIs;

Fig. 7 illustrates use of extended URIs for fragment location according to the preferred embodiment;and

Fig. 8 is a schematic block diagram of a general purpose computer upon which the preferred embodiment of the present invention can be practiced.

Detailed Description including Best Mode

Where reference is made in any one or more of the accompanying drawings to steps and/or features, which have the same reference numerals, those steps and/or features

have for the purposes of this description the same function(s) or operation(s), unless the contrary intention appears.

In the context of this specification, the word "comprising" means "including principally but not necessarily solely" or "having" or "including" and not "consisting only of". Variations of the word comprising, such as "comprise" and "comprises" have corresponding meanings.

It is noted that the introductory part of the description makes reference, for illustrative purposes, to audio and video content which is stored on Compact Disk Read Only Memory (CD ROM) media accessed by a "Juke Box" device which is capable of storing a number of such disks and accessing them according to an address.

Fig. 1 depicts a prior art system used to locate an audio content CD ROM 312 using the Internet 308 as a vehicle. A user (not shown) uses a personal computer (PC) 304 which is connected to the Internet 308 in order to connect to the server 306 of an on-line music provider. The server 306 is connected to a CD ROM juke box 310 which houses a plurality of CD ROMs 312, 316. Each CD ROM 312, 316 contains individual songs exemplified, for illustrative purposes only, by bold lines 314, 318 respectively. The user has a paper description 300 of the desired CD ROM 312 containing a title 326 of the CD ROM, and also a list of the songs 302, 320. The user uses a Universal Resource Indicator (URI) 324 which "points" to the address of the CD ROM 312, and the user is able to download music from the CD ROM 312 over the system.

In Fig. 2, the CD ROM 312 can be portrayed in a description 400 as containing a list of songs 402, 404 under a title 414, where the song 404 has indices 410 and 412 which point to particular segments within the song 404. Terminology such as "songs" is used for illustration in this part of the description, noting that in fact, as described, the audio content is actually stored on CD ROM as noted. For example in classical music

where the "song" 402 can be an individual movement of a symphony, and therefore can be quite long, the index 1 (ie 410) can point to a trumpet solo, and the index r (ie 412) can point to a violin solo of interest. Depending upon the capabilities of the server 306 and juke box 310 in Fig. 1, the user can address the desired CD ROM 312, and address a desired index 410. It is noted however, with reference to both Fig. 1, and to Fig. 2, that the user is limited to addressing, and so accessing, material only down to the level of the particular CD (ie 312), or perhaps the specified predefined index (ie 410). It is not possible to "drill down" to an arbitrary further specified level of fine grain detail.

Fig. 3 depicts an illustrative embodiment of an addressing method, in this case to be used in relation to audio CD ROMs. The CD ROM 312, formerly described by the description 400 which contains a list of individual songs 402, each of which may contain a level of indexing (eg 410 see Fig. 2) is extended, using a logical model based upon consecutive time blocks or slices, into a hierarchical representation comprising both the description 400 and the further decription 500 comprising time blocks 502 to 512. The logical model, when applied to the CD ROM 312, serves to form a hierarchical representation of the otherwise monolithic AV content of the CD ROM 312. The model thereby enables systematic and rapid addressing of arbitrary content fragments on a time block basis, and provides the desired arbitrary drill down capability. Using the described representation, a user is able, for example, to select an arbitrary fragment of audio content on the CD ROM 312 by specifying a fragment address, or fragment identifier, of the form "Title / song1 / block 2 – block j-3", where j is an arbitrary index as shown in Fig. 3. The present logical model is used for illustrative purposes, and more advantageous logical models and addressing schemes are proposed later in the description. In Fig. 3 song 402 is shown to comprise blocks 502 to 506, song 2 comprising blocks from the block after 504 through to block 506 and so on.

Turning to Fig. 4, the hierarchical representation 602, comprising descriptions 400 and 500, is shown in a system content in more detail. By using the fragment identifier 610, derived from the hierarchical representation 602, in conjunction with the URI 324 (see Fig. 1), an extended URI 606 (hereafter referred to as a "URI reference" is shown to incorporate both the URI 324 described in relation to Fig. 1, and an additional fragment identifier 608. The URI reference can thus be used as an address to the CD 312, and further, to the desired fragment 314.

Fig. 5 depicts another hierarchical representation 706, determined using a logical model appropriate for digital video. In this example, a sequence of digital video shots 700 is recorded on a CD ROM 724. The logical model selected resolves the video sequence 700 into frames eg 708, each frame being further resolved into x intervals eg 710 and y intervals eg 722. This logical model is used for illustrative purposes, and more advantageous logical models are proposed later in the description. Using the described representation, a user is able, for example, to select an arbitrary spatial fragment of video content on a specified frame of the CD ROM 724 by specifying a fragment address, or fragment identifier, of the form "Title / frame1 / x1 – x2; y1 – y2".The x interval from x1 (726) to x2 (728) and the y interval from y1 (730) to y2 (732) address the spatial region 704 within the frame 702 in the set of digital video shots 700. The URI reference 716 therefore contains a portion 734 prior to the hash sign 720 which addresses the digital video disc 724, while the portion 736 after the hash sign 720 addresses the fragment 704.

Having provided an illustrative description of an embodiment of the invention, a more detailed description is now provided. XML is utilised as a basis for describing a preferred embodiment of the present invention. This is both from the standpoint of conceptual and notational convenience, and also because XML has significant support as a recommendation in the context of the World Wide Web Consortium (W^3C).

It is shown, in the preferred embodiment, how the XML Linking Language XLink, the XML Pointer Language XPointer, and the [illegible] XPath, can be extended in order to locate fragments of non XML-based audio-visual content.

XLink uses URI's for locating objects. In principle, modified URIs can be used for locating any resource that has identity, for instance, an electronic document, an image, a service, a collection of other resources, a person, an corporation, or a bound book in a library. Each resource corresponds to an entity or set of entities in a conceptual model. URI's can therefore be used for locating or referencing resources other than XML documents. However, the XPath and XPointer schemes that XLink currently uses for addressing the internal structure of data objects can only be used to locate fragments of XML documents.

As an introduction, the use of XLink, XPointer, and XPath, are considered in the limited context of XML documents. XPath models an XML document as a tree of nodes. There are seven types of nodes, namely root nodes, element nodes, text nodes, attribute nodes, namespace nodes, processing instruction nodes and comment nodes. XPath uses a compact, non-XML syntax to facilitate the use of XPath within URI's. An XPath location path consists of a '/'-separated list of location steps. Each location step has the form:

axis :: node-test [predicates]

where axis specifies the tree relationship between the nodes selected by the location step and the context node; node-test specifies the node type or the name; and predicates refine the set of nodes selected by the location step.

A number of syntactic abbreviations allow common cases to be expressed concisely as follows:

@ is short for attribute::, e.g. attribute::type can be abbreviated as @type,

// is short for /descendant-or-self::node()/,

. is short for self::node(), and

.. is short for parent::node().

An axis specifies the tree relationship between the nodes selected by the location step and the context node. XPath axes include *child*, *parent*, *descendant*, *ancestor*, *following-sibling*, *preceding-sibling*, *following*, *preceding*, *attribute*, *namespace*, *self*, *descendant-or-self* and *ancestor-or-self*. The default is the *child* axis. XPointer extends XPath adding the *string* and *range* axes.

A node test specifies the node type or the name (such as the name of an element or an attribute) of the nodes selected by the location step.

There can be zero or more predicates for refining the set of nodes selected by the location step. Predicates are evaluated for each candidate location along the specified axis, and typically test the element type, attributes, positions, and/or other properties of the candidate nodes.

A function library provides a set of predicate functions such as count(), position(), id(), last(), etc. Each function takes zero or more arguments and returns a single result. Like XPointer, a new scheme can define new functions to extend the core functions of XPath.

Each location step is evaluated with respect to a context. The context is initially the document root, or more generally the results of a prior location step. The node set selected by the location step is the node set that results from generating an initial node set from the axis and node test, and then filtering that node-set by each of the predicates in turn.

Some examples of XPath location paths are as follows:

/doc/chapter[2]/section[3]

selects the third section of the second chapter of doc

chapter[contains(string(title), "Overview"]

selects the chapter children of the context node that have one or more title children containing the text "Overview"

child::*[self::appendix or self::index]

selects the appendix and index children of the context node

child::*[self::chapter or self::appendix] [position()=first()]

selects the first chapter and appendix children of the context node

para[@type="warning"]

selects all para children of the context node that have a type attribute with value "warning"

para[@id]

selects all the para children of the context node that have an id attribute.

XPath operates on the abstract, logical structure of an XML document. For instance, the examples given in the previous section assumed an XML document with the structure, in Extended Backus-Naur Form (ENBF) as follows:

```
doc ::= toc chapter+ appendix* index
chapter ::= section+
section ::= para+
appendix ::= section+
```

where "toc" means "table of contents", "+" means "one or more, "*" means zero or more", and the composite description presented above describes, in expanded form, a document comprising a table of contents, one or more chapters, zero or more appendices,

and an index, where each chapter comprises one or more sections, where each section comprises one or more paragraphs, and finally where each appendix comprises one or more sections.

In an XML document, each of these structures is marked by a pair of appropriately named tags. The tag markup allows the logical structure of the document to be determined unambiguously. Hence, any application that understands the syntax of XML can determine the location of the document's components. Any application that understands the XPath and XPointer notations can use an URI with an XPath/XPointer fragment identifier to locate parts of the document.

Audio-visual, or AV content, is not stored as XML documents and cannot be marked up. However, given an unambiguous logical structure, or model, a modified XPath location/addressing method can be used. Hence, for each class of AV content, in the first instance, an unambiguous logical structure must be defined. By an unambiguous logical structure or model, it is meant that different persons and applications will segment given content in exactly the same way given the model.

Considering one type of AV content, for instance, Digital Video format as used by digital video cameras, this can be modelled as:

dv ::= frame*

where this means a digital video comprising one or more frames.

In the case, for example, where compatible digital video cameras generate meta-data to represent and record the instances the camera starts recording (designated a REC event), a *shot* can be defined as an interval between two REC event. In this case, the model for DV format is:

dv ::= shot*

shot ::= frame+

meaning a digital video comprising zero or more shots, each shot comprising one or more frames.

As another example, the logical structure of Compact Disc Audio can be modelled as follows:

cdAudio ::= track*

track ::= channel channel index*

channel ::= sample*

meaning an audio CD comprising zero or more tracks, each track comprising two channels, and zero or more indices, and each channel comprising zero or more samples.

Considering a more complex example, consider Digital Video Disc, or DVD, video which can provide:

over 2 hours of high-quality digital video (over 8 on a double-sided, dual-layer disc),

up to 8 tracks of digital audio, each with as many as 8 channels

up to 32 subtitle/karaoke tracks

up to 9 camera angles (different viewpoints) can be selected during playback

up to 32 separate subpicture channels

Other data types include Video Manager Information files, Video Title Set files, Program Chain Information files, still picture Video Objects, attributes for Title, Part_of_Titles, and Menus, Time Map Tables, Part_of_Title Search Pointers, and Navigation Commands.

DVD-Video content is broken into *titles* and *chapters* (or parts of titles). Titles are made up of *cells* linked together by one or more *program chains* (PGC). Individual cells can be used by more than one PGC. Different PGCs define different sequences through mostly the same material. Additional material for camera angles and branching is interleaved together in small chunks. The DVD player jumps from chunk to chunk, skipping over unused angles or branches, to stitch together the seamless video.

One logical model for DVD-Video is:

dvdVideo ::= mainMenu? title* subpicture* file*

mainMenu ::= menu*

menu ::= menu*

title ::= chapter+

chapter ::= view+ audio+ subtitle*

view ::= frame+

audio ::= channel+

channel ::= sample+

The previous logical models each relate to a class of AV content, namely digital video, compact disc audio, and digital video disk. As noted, the application of the logical models to the associated AV content produces hiearchical representations of the AV content which supports addressing of fragments of the content.

Turning to the aspect of addressing, each location step is evaluated with respect to a context. The context is initially the root node, dvdVideo in this case. In general, the context is the results of a prior location step. The node set selected by the location step is the node set that results from generating an initial node set from the axis and node test, and then filtering that node-set by each of the predicates in turn.

In the above example, it works as follows:

	Selection step	Meaning	Context after the selection step
axis	(default is child::)	all the children of the context node	all the children of the dvdVideo node, that is, all thc mainMcnu, title, subpicture and file nodes
node-test and predicate	Title[2]	any title node whose position is 2	the 2nd title nodes
axis	Time::	arrange the current selection continuously in time starting at time zero	unchanged
node-test	0m,15m	all content inside the time interval 0min to 15min	the first 15 minutes of the 2nd title

As another example,

http://www.apxcom.com/products/dvd0111#avptr(/dvdVideo/title[(position()=1 or position()=2][time("0m","15m")]

selects the first 15 minutes of the first and the second titles of the DVD.

It works as follows:

	Selection step	Meaning	Context after the selection step
axis	(default is child::)	all the children of the context node	All the children of the dvdVideo node, that is, all the mainMenu, title, subpicture and file nodes
node-test and predicate	title[position()=1 or position()=2]	any title node whose position is either 1 or 2	The first and the second title nodes
additional predicate	[time("0m","15m")]	any content inside the time interval 0min to 15min (of each candidate node)	the first 15 minutes of the first and the second titles

As another example,

http://www.apxcom.com/products/acd010239#avptr(/cdAudio/track[2]/channel/time::0s,60s)

selects the first minute of the second track of an audio CD which has the model

cdAudio ::= track*

track ::= channel channel index*

channel ::= sample*

The location steps work as follows:

	Selection step	Meaning	Context after the selection step
axis	(default is child::)	all the children of the context node	all the children of the cdAudio node, that is, all the track nodes
node-test and predicate	track[2]	any track node whose position is 2	the 2^{nd} track node
axis	(default is child::)	all the children of the context node	all the children of the 2^{nd} track node
node-test and predicate	channel	any channel node	all the channel nodes of the 2^{nd} track node
axis	time::	arrange the current selection continuously in time starting at time zero	unchanged
node-test	time::0s,60s	all content inside the time interval 0sec to 60sec	the first minute of the 2^{nd} track

The described AV location scheme can, utilising a notation and mechanism similar to those of XPath/Xpointer, locate analog and digital AV content within a database, or a plurality of databases.

A set of named functions are defined for the AV location scheme. For instance,

time(startTime [, endTime])	For determining whether the current context is within the specified time.
timecode(startTimecode [, endTimecode])	For determining whether the current context is within the time specified by the start and end timecodes.

The functions can be used for evaluating expressions, the evaluation always occurring with respect to the current context.

In addition, new axes can be added, for instance, a time axis and a region axis for locating temporal and spatial segments of the data. The incorporation of these axes provides additional power to the concept of fragment addressing, and allows drilling down to different aspects of the AV content.

The time axis selects, within the current context, components that occur within the specified start and end time. The current context is taken as starting at time zero and progressing continuously through time in normal play time. If the end time is not specified, it is taken to be the same as the start time and the component that occurs at or closest to the specified start time is selected.

```
TimeLocationStep ::= 'time' '::' StartTime ( ',' EndTime )?
TimeUnit ::= 'h' | 'm' | 's' | 'ms'
TimeNotation ::= 'end' | ( [0-9]+ TimeUnit )
StartTime ::= TimeNotation
EndTime ::= TimeNotation
```

For example,

http://www.apxcom.com/products/dvd0111#avptr(/dvdVideo/title[2]/ time::0m,15m)

selects the first 15 minutes the second title of the specified DVD

http://www.apxcom.com/products/dvd0111#avptr(/dvdVideo/title[position()=1 or position()=2)][time("0m","15m")]

selects the first 15 minutes of the 1st and the 2nd titles of the specified DVD

The timecode axis selects, within the current context, components that occur within the specified start and end timecode. If the end timecode is not specified, it is taken

to be the same as the start timecode and the component that occurs at the specified start timecode is selected. The timecode is represented by a time value or a combination of date and time values as defined in International Standards Organisation (ISO) 8601. It can also be an SMPTE (ie Society of Motion Picture and Television Engineers) timecode in the format of HH:MM:SS:FF where FF stands for frame.

```
TimecodeLocationStep ::= 'timecode' '::' StartTimecode ( ',' EndTimecode )?
TimecodeNotation ::= 'begin' | 'end' | smpteTimecode | time | ( '"' dateTime '"' )
StartTimecode ::= TimecodeNotation
EndTimecode ::= TimecodeNotation
```

For example,

http://www.apxcom.com/events/productLaunch99#avptr(/dv/shot/timecode::00:15:00:00,00:30:00:00)

selects 15 minutes of clips on the specified digital video tape using SMPTE timecodes

The region axis selects, within the current context, the 2D region that is bounded by the specified bounding curve. The origin corresponds to the top-left corner of a frame with the x- and y-axis coordinates increasing to the right and down. Coordinates are specified in (integral) pixel values. Several types of bounding curves such as rectangle and ellipse allow the (anti-clockwise) angle between its major axis and the x-axis to be specified. To allow a region to be specified for different resolutions of the same content, the resolution of the source from which the bounding curve is determined could be specified using the range() function.

```
RegionLocationStep ::= 'region' '::' [ Range ] BoundingCurve
BoundingCurve ::= Shape
Shape ::= Circle | Ellipse | Rectangle | Polygon | QuadCurve | CubicCurve | Bspline
Circle ::= 'circle(' XCentre ',' YCentre ',' Radius ')'
Ellipse ::= 'ellipse(' XCentre ',' YCentre ',' Major ',' Minor ',' Angle ')'
Rectangle ::= 'rect(' Left ',' Top ',' Width ',' Height ',' Angle ')'
```

Polygon ::= 'polygon(' Point ',' Point (',' Point)+ ')'

QuadCurve ::= 'qcurve(' Point ',' Point (',' Point)+ ')'

CubicCurve ::= 'ccurve(' Point ',' Point (',' Point)+ ')'

BSpline ::= 'bspline(' Point ',' Point (',' Point)+ ')'

Point ::= Integer ',' Integer

Integer ::= [+|-] Digits

Angle ::= Degree

Range ::= 'range(' Integer ',' Integer ')'

For example,

http://www.apxcom.com/events/productLaunch99#avptr(/dv/shot[1]
/frame[1012]/region::range(720,480)rect(40,40,60,60,45)

selects a 60x60 diamond-shape region from the 1012th frame of shot one of the specified digital video tape

Using the models and structures described above, a URI reference for the first minute of the second track of an audio CD can have the form:

http://www.apxcom.com/products/acd010239#avptr(/cdAudio/track[2]/channel/time::0s,60s)

where the portion of the URI reference before the hash refers to the AV product, in this instance an audio CD no. 010239, belonging to a class of products, ie audio CDs, produced by the fictitious company "apxcom", which is the resource in this instance. What follows the hash sign is an "AV" fragment identifier, or pointer for locating specific AV content on the designated CD. The AV pointer is directed to the internal structure of the content, in this case the first minute of the second track. Therefore, the URI (Universal Resource Indicator), which is the familiar entity used in Internet addressing, when combined with a fragment identifier (the part of the URI reference following the hash sign), is called a URI reference.

In a further example, the URI to a 15 minutes segment of the second movie on a DVD can have the form:

http://www.apxcom.com/products/dvd0111#avptr(/dvdVideo/title[2]/time::30m,15m)

while

http://www.apxcom.com/products/dvd0111#avptr(/dvdVideo/title/chapter/audio[1])

will select the first audio track of the DVD.

In considering the above URI references, it is again noted that the portion of the URI before the hash sign refers to the AV product, namely the resource. What follows the hash sign is an AV fragment identifier for locating parts of the AV content.

Fig. 6 presents a description of a prior art scenario in the context of an XML document 114 presented on a browser (not shown). It is noted that the document 114 as shown depicts a physical aspect of the XML document, whereas a user of the browser would be presented with the document in a different style (not shown). The document 114 describes AV content about the Apollo 13 space mission. The document type is designated by a reference 100, which in the present instance is a "Documentary". A number of hyperlink references 104, 106 provide links to other XML documents 120, 122, which describe movie sources and movie reviews (not shown) respectively. The movie sources referred to by document 120 can be either on-line, or alternatively, can be a physical entities such as a video-casette 128 (depicting a specific movie source in this example) produced by a company. The document 114 contains a segment 110 named "Rocket Launch" between tag delimiters 108 and 124. The rocket launch segment commences 15 minutes after the start of the documentary as indicated by the start index 112. When a user selects the release reference 104, the associated link depicted by an arrow 118 directs the user to the XML document 120 describing the movie source 128 as already noted. Selection of a reference 132 on the document 120 retrieves a URI 126, which points, as indicated by an arrow 130, to the physical video cassette 128. The URI 126 is seen to comprise a link to a company having a domain name movies designated 134, where the specific casette is designated "vhs0111" ie 136 in the "products" category

138. Reviewing the aforementioned process for clarity of explanation, selection by the user of the Apollo 13 reference 104 directs the user to the specific cassette 128 through a location process depicted by the URI 126. Noting that the URI has a standard format and does not incorporate a fragment identifier, it is clear that the described scenario does not support drilling down to the fragment level. The URI does address the particular video casette 128, but provides no mechanism for addressing AV data on the casette at the fragment level.

Turning to Fig. 7, a preferred embodiment of the proposed addressing method is described. Selection of the preview reference 200 on the main XML document 114 activates a link depicted by an arrow 202 which is directed to another XML document 204. This latter document 204 describes preview AV material at the fragment level, and selection of a reference 226 results in a link depicted by a dashed arrow 206 pointing to an AV fragment using an extended URI 208. The portion of the URI 208 before the hash relates to VHS Preview content designated "01100", which is a product of a fictitious company called "movies". The portion of the URI after the hash is an AV pointer 212, pointing to the second video 216 of the vhs (214) tape, and in particular to a segment (218) starting 900 seconds after the start of the documentary, and ending 1800 seconds after the start of the documentary.

The method of addressing an arbitrary fragment of an AV resource is preferably practiced using a conventional general-purpose computer system 800, such as that shown in Fig. 8 wherein the processes of Figs. 3 to 5 and 7 may be implemented as software, such as an application program executing within the computer system 800. In particular, the steps of the method of addressing an arbitrary fragment of an AV resource are effected by instructions in the software that are carried out by the computer. The software may be divided into two separate parts; one part for carrying out the addressing methods;

and another part to manage the user interface between the latter and the user. The software may be stored in a computer readable medium, including the storage devices described below, for example. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable medium having such software or computer program recorded on it is a computer program product. The use of the computer program product in the computer preferably effects an advantageous apparatus for addressing an arbitrary fragment of an AV resource in accordance with the embodiments of the invention.

The computer system 800 comprises a computer module 801, input devices such as a keyboard 802 and mouse 803, output devices including a printer 815 and a display device 814. A Modulator-Demodulator (Modem) transceiver device 816 is used by the computer module 801 for communicating to and from a communications network 820, for example connectable via a telephone line 821 or other functional medium. The modem 816 can be used to obtain access to the Internet, and other network systems, such as a Local Area Network (LAN) or a Wide Area Network (WAN).

The computer module 801 typically includes at least one processor unit 805, a memory unit 806, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 807, and an I/O interface 813 for the keyboard 802 and mouse 803 and optionally a joystick (not illustrated), and an interface 808 for the modem 816. A storage device 809 is provided and typically includes a hard disk drive 810 and a floppy disk drive 811. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 812 is typically provided as a non-volatile source of data. The components 805 to 813 of the computer module 801, typically communicate via an interconnected bus 804 and in a manner which results in a conventional mode of operation of the computer

system 800 known to those in the relevant art. Examples of computers on which the embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

Typically, the application program of the preferred embodiment is resident on the hard disk drive 810 and read and controlled in its execution by the processor 805. Intermediate storage of the program and any data fetched from the network 820 may be accomplished using the semiconductor memory 806, possibly in concert with the hard disk drive 810. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 812 or 811, or alternatively may be read by the user from the network 820 via the modem device 816. Still further, the software can also be loaded into the computer system 800 from other computer readable medium including magnetic tape, a ROM or integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 801 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including email transmissions and information recorded on websites and the like. The foregoing is merely exemplary of relevant computer readable mediums. Other computer readable mediums may be practiced without departing from the scope and spirit of the invention.

The method of addressing an arbitrary fragment of an AV resource may alternatively be implemented in dedicated hardware such as one or more integrated circuits performing the functions or sub functions of addressing. Such dedicated hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories.

Adoption of XML as a notation for describing the preferred embodiment is a convenient mechanism for describing the embodiment. It also allows a consistent view

and addressing mechanism for both XML and non-XML resources. As noted previously however, this is not an essential feature of the present invention.

Industrial Applicability

It is apparent from the above that the embodiments of the invention are applicable to the computer and data processing industries.

The foregoing describes only some embodiment of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the invention, the embodiments being illustrative and not restrictive.

Claims:

1. A method for addressing an arbitrary fragment of an audio-visual (AV) resource belonging to a class of AV resources, the method comprising the steps of:

identifying a logical model for the class of AV resources;

applying the model to the AV resource to form a hierarchical representation of said AV resource including a representation of the AV fragment;

determining a first address for the AV resource;

determining a second address for the representation of the AV fragment depending upon the hierarchical representation; and

combining the first and second addresses to determine an address for the AV fragment.

2. A system for addressing an arbitrary fragment of an audio-visual (AV) resource belonging to a class of AV resources, the system comprising:

identification means for identifying a logical model for the class of AV resources;

application means for applying the model to the AV resource to form a hierarchical representation of said AV resource said AV resource representation including an associated root node and a representation of the AV fragment;

first determination means for determining a first address for the AV resource root node;

second determination means for determining a second address for the representation of the AV fragment depending upon the hierarchical representation; and

combining means for combining the first and second addresses to determine an address for the AV fragment.

3. A system according to claim 2, wherein the first address is a Universal Resource Identifier (URI), the second address is a fragment identifier, and wherein the address determined by combining the URI and the fragment identifier is a URI reference.

4. A method for addressing an arbitrary fragment of an audio-visual (AV) resource belonging to a class of AV resources, the method comprising the steps of:

determining a first address for the AV resource; characterised in that the method identifies a logical model for the class of AV resources, whereby applying the model to the AV resource forms a hierarchical representation of said AV resource including a representation of the AV fragment, the method comprising the further steps of;

determining a second address for the representation of the AV fragment depending upon the hierarchical representation; and

combining the first and second addresses to determine an address for the AV fragment.

5. A method for addressing an arbitrary fragment of an audio-visual (AV) data set, whereby application of a first logical model to the AV data set according to a first set of rules has formed a one-to-one meta-data representation of the AV data set, said representation of the AV data set including at least a meta-data representation of said fragment, the meta-data representation of the fragment being associated with a reference, said method comprising the steps of:

selecting the reference associated with the meta-data representation of the fragment; and

applying a second logical model to the selected reference according to a second set of rules to form a meta-data path pointing to the fragment.

6. A system for addressing an arbitrary fragment of an audio-visual (AV) data set, said system comprising:

first application means for applying a first logical model to the AV data set according to a first set of rules to form a one-to-one meta-data representation of the AV data set, said representation of the AV data set including at least a meta-data representation of said fragment, said meta-data representation of the fragment being associated with a reference;

selection means for selecting the reference associated with said meta-data representation of the fragment; and

second application means for applying a second logical model to the selected reference according to a second set of rules to form a meta-data path pointing to the fragment.

7. A computer program product including a computer readable medium having recorded thereon a computer program for implementing a method for addressing an arbitrary fragment of an audio-visual resource, said product comprising: code for identifying a logical model for the class of AV resources;

code for applying the model to the AV resource to form a hierarchical representation of said AV resource including a representation of the AV fragment;

code for determining a first address for the AV resourcc;

code for determining a second address for the representation of the AV fragment depending upon the hierarchical representation; and

code for dcombining the first and second addresses to determinc an address for the AV fragment.

DATED this Twenty-seventh Day of September 1999

Canon Kabushiki Kaisha

Patent Attorneys for the Applicant

SPRUSON & FERGUSON

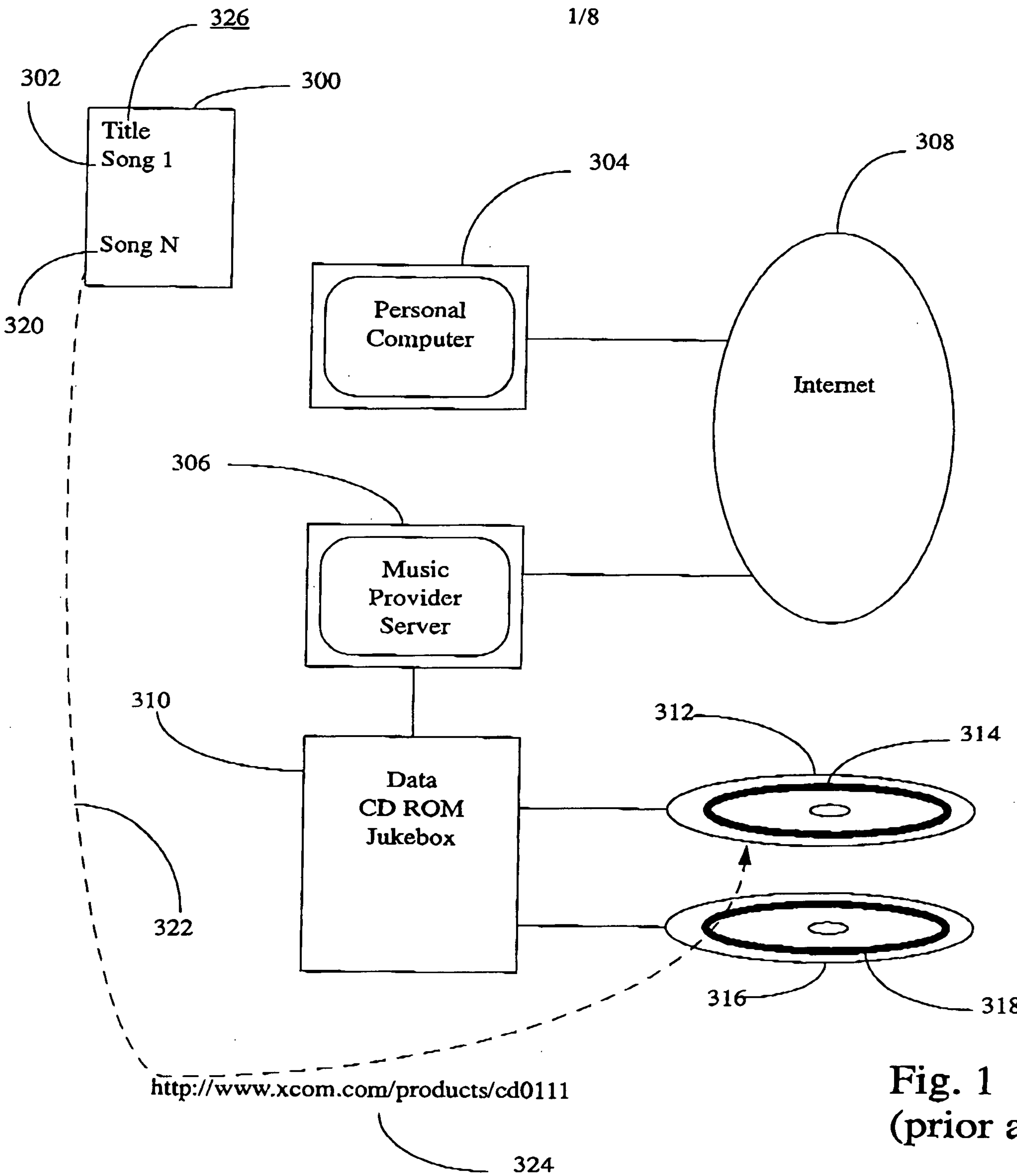


Fig. 1
(prior art)

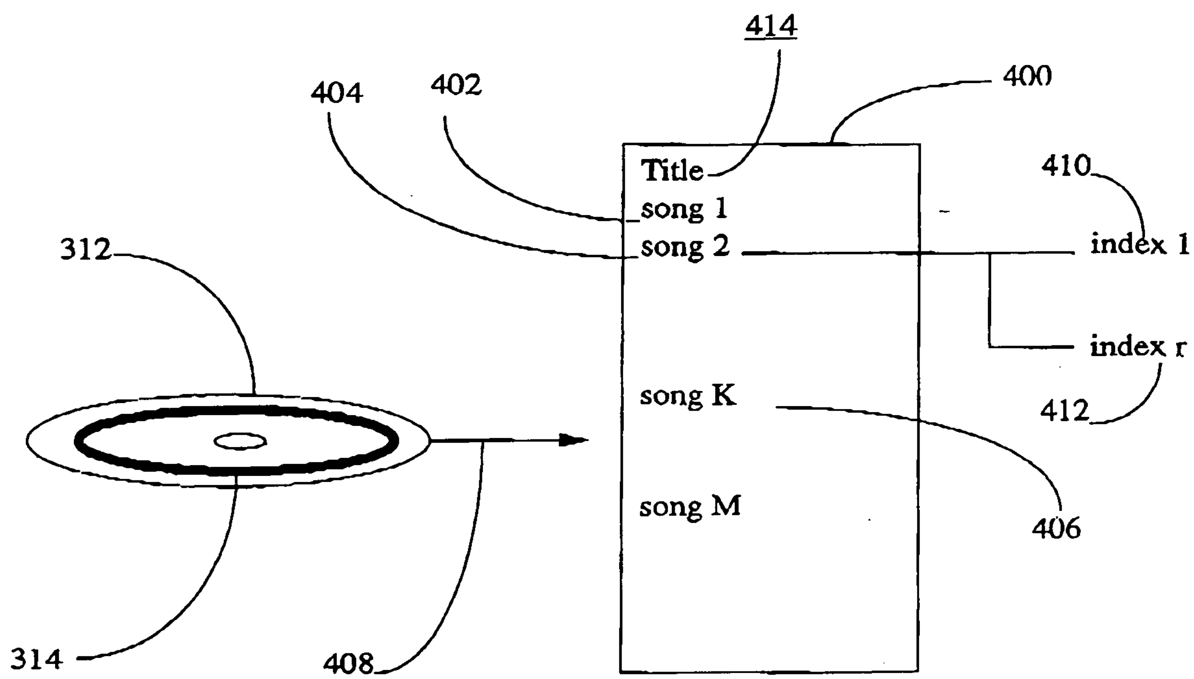
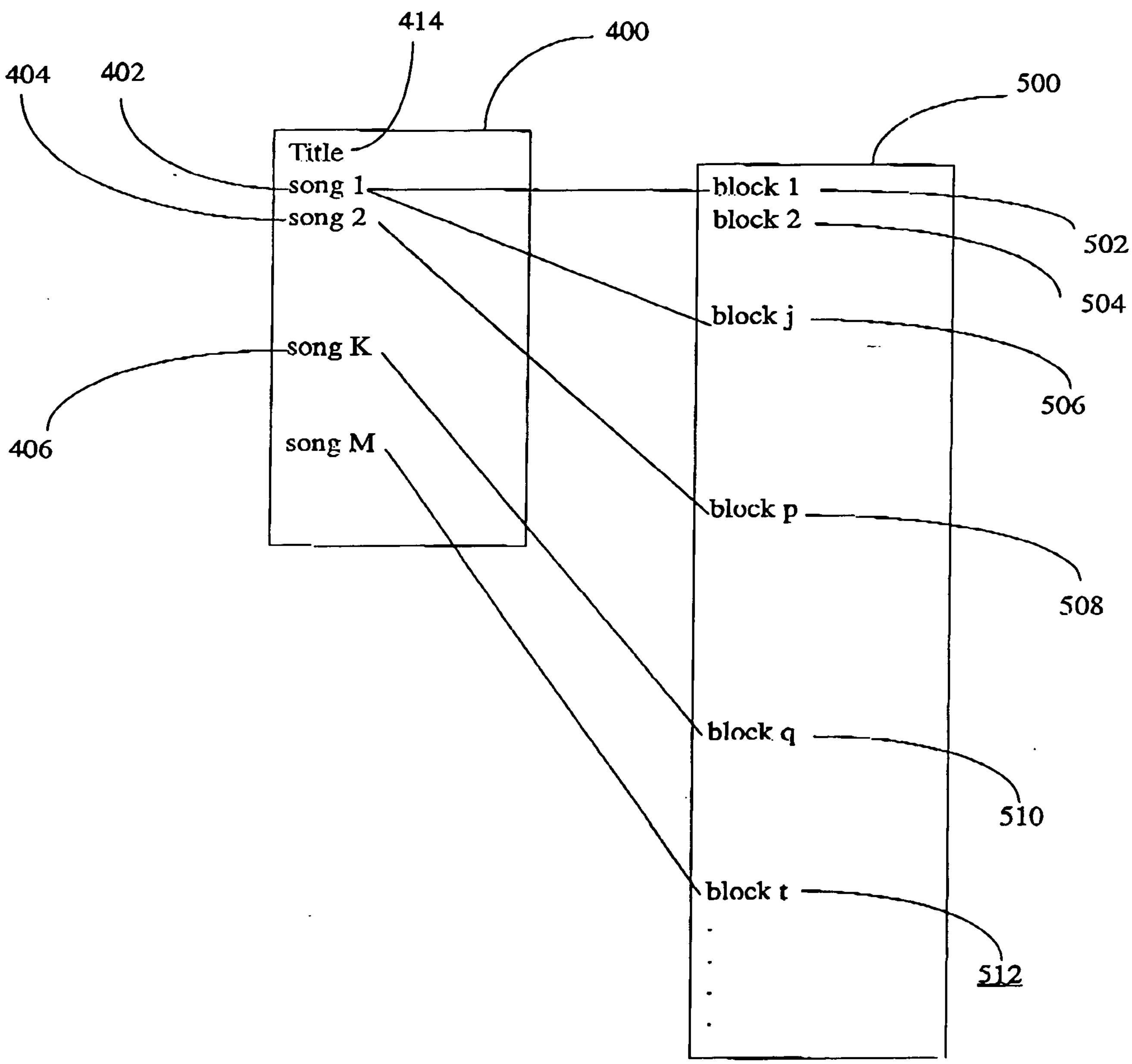


Fig. 2
(prior art)



414
400
404
402
500
Title
song 1
song 2
song K
song M
block 1
block 2
block j
block p
block q
block t
502
504
506
406
508
510
512

Fig. 3

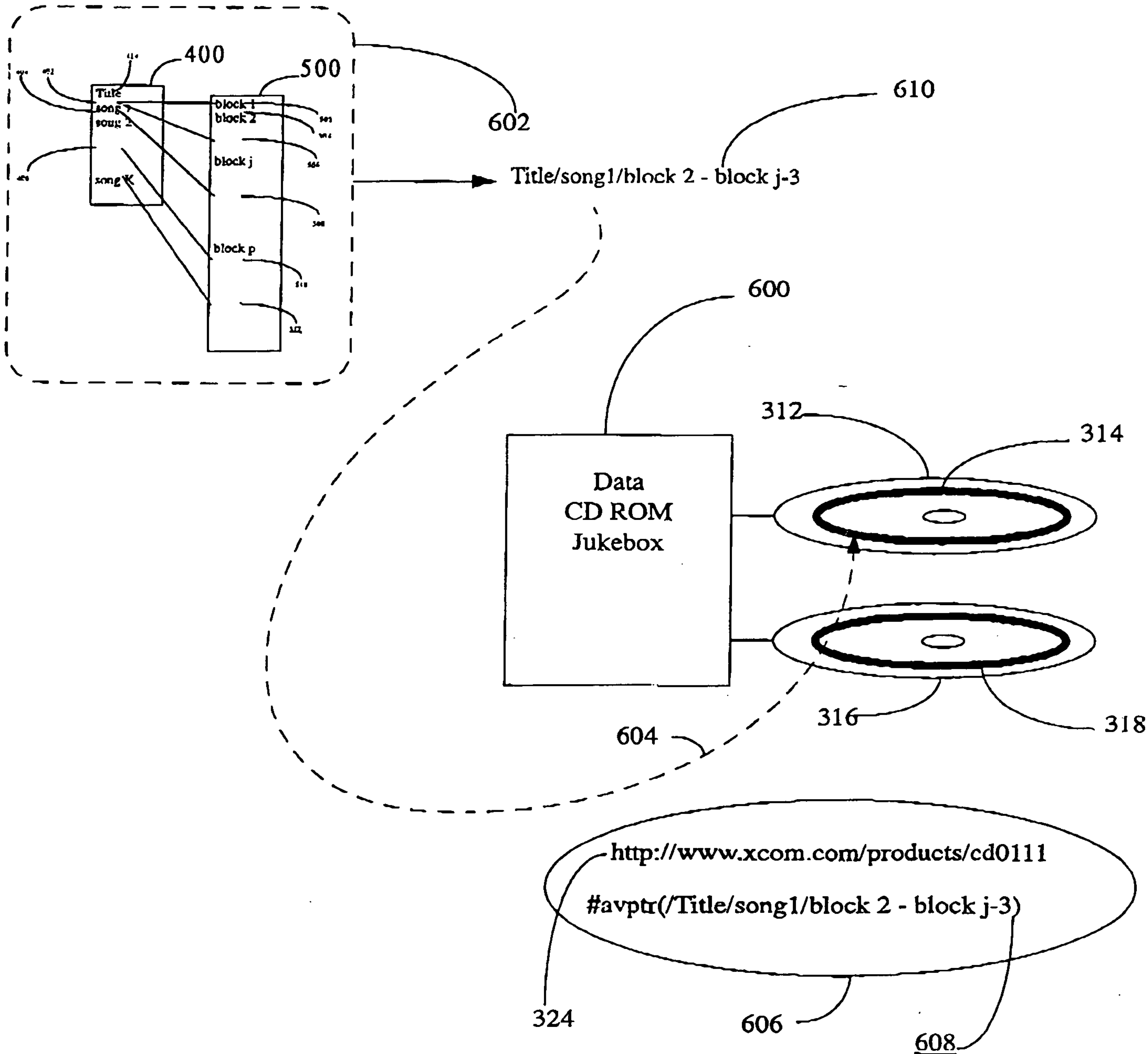

400
500
Title
song 1
song 2
song K
block 1
block 2
block j
block p
602
610
Title/song1/block 2 - block j-3
600
Data
CD ROM
Jukebox
312
314
316
318
604
http://www.xcom.com/products/cd0111
#avptr(/Title/song1/block 2 - block j-3)
324
606
608

Fig. 4

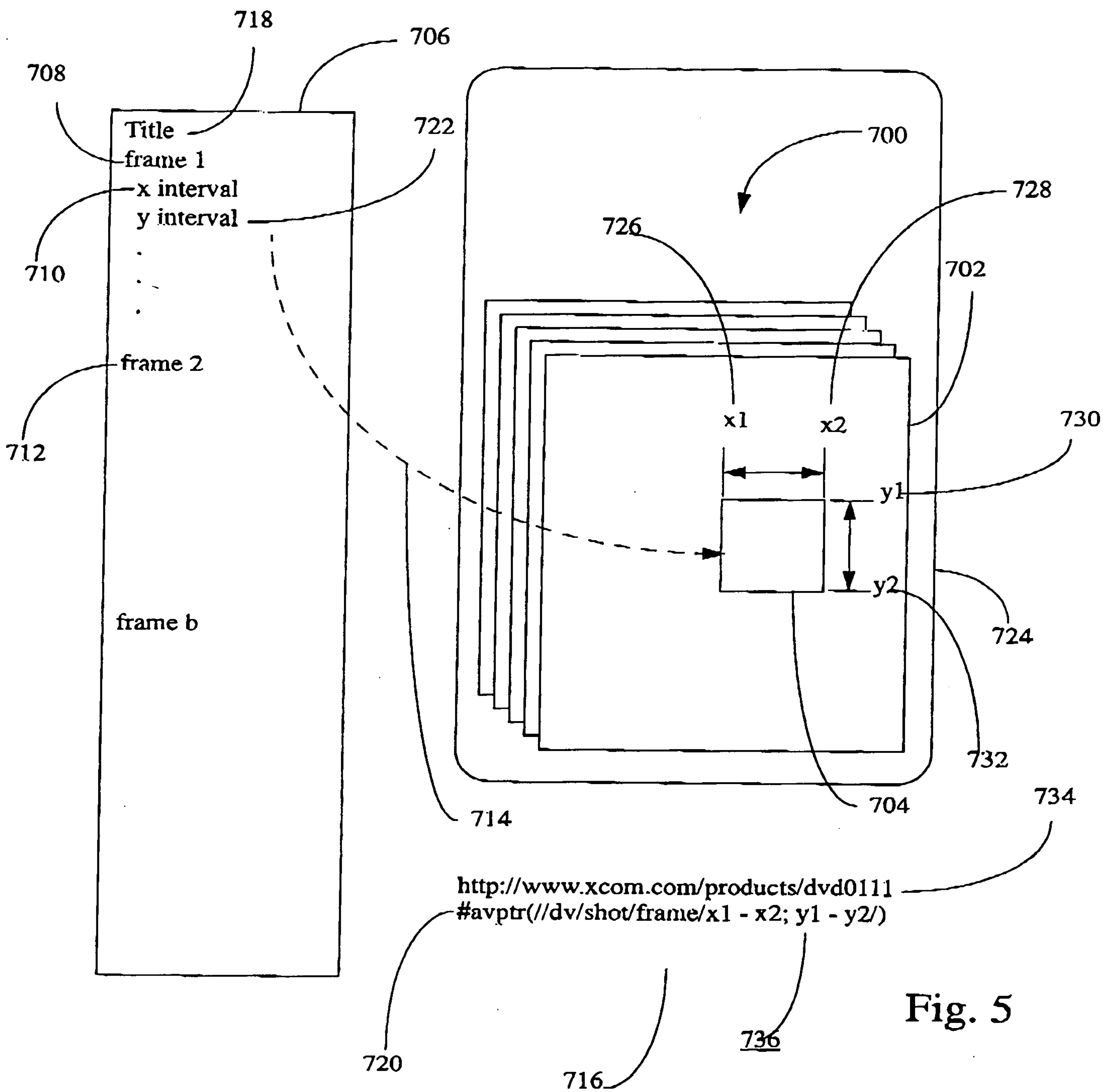

718
706
708
Title
frame 1
x interval
y interval
722
710
frame 2
712
frame b
700
728
726
702
x1
x2
730
y1
y2
724
732
714
704
734
http://www.xcom.com/products/dvd0111
#avptr(//dv/shot/frame/x1 - x2; y1 - y2/)
720
716
736

Fig. 5

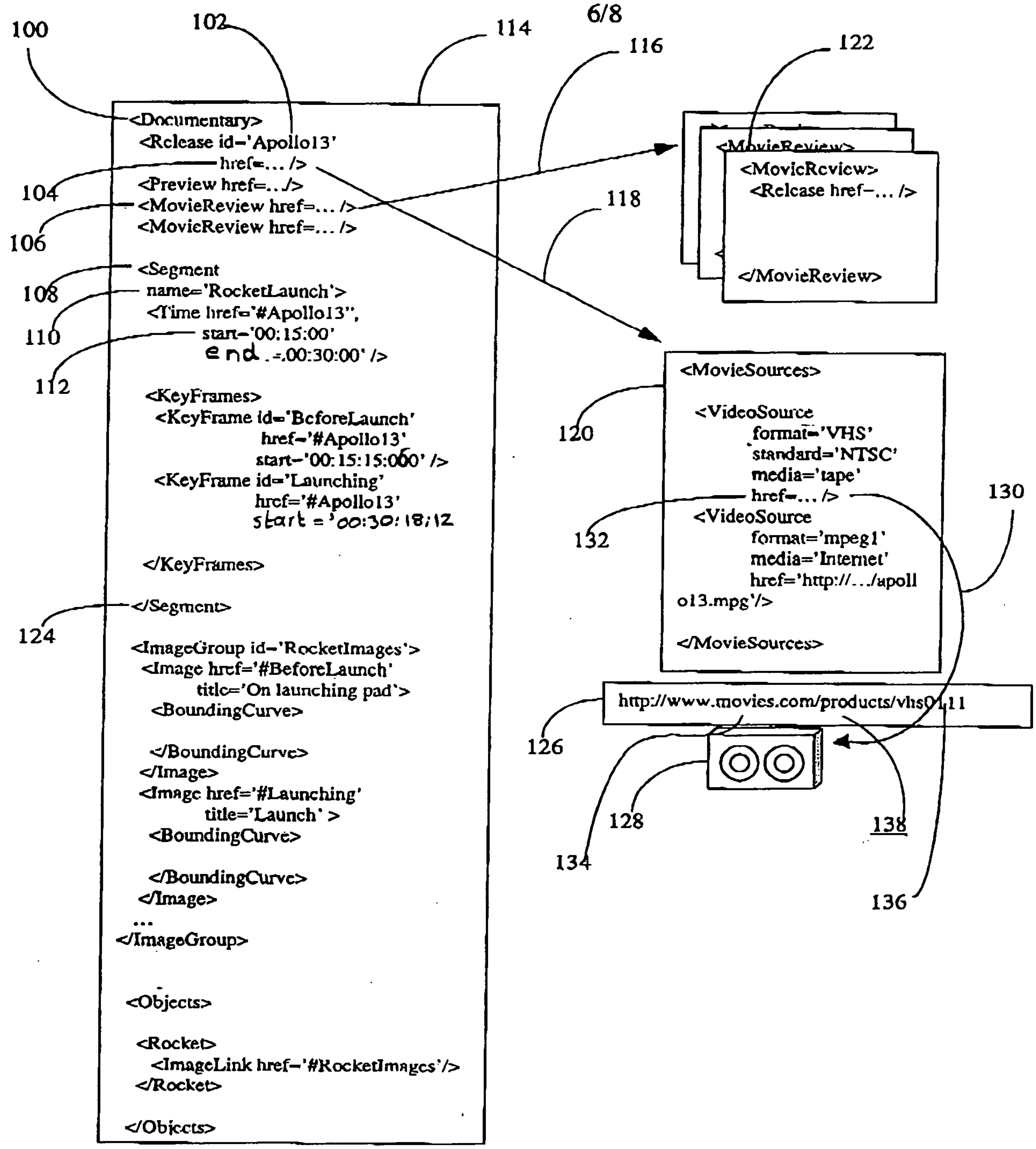


Fig. 6
(Prior Art)

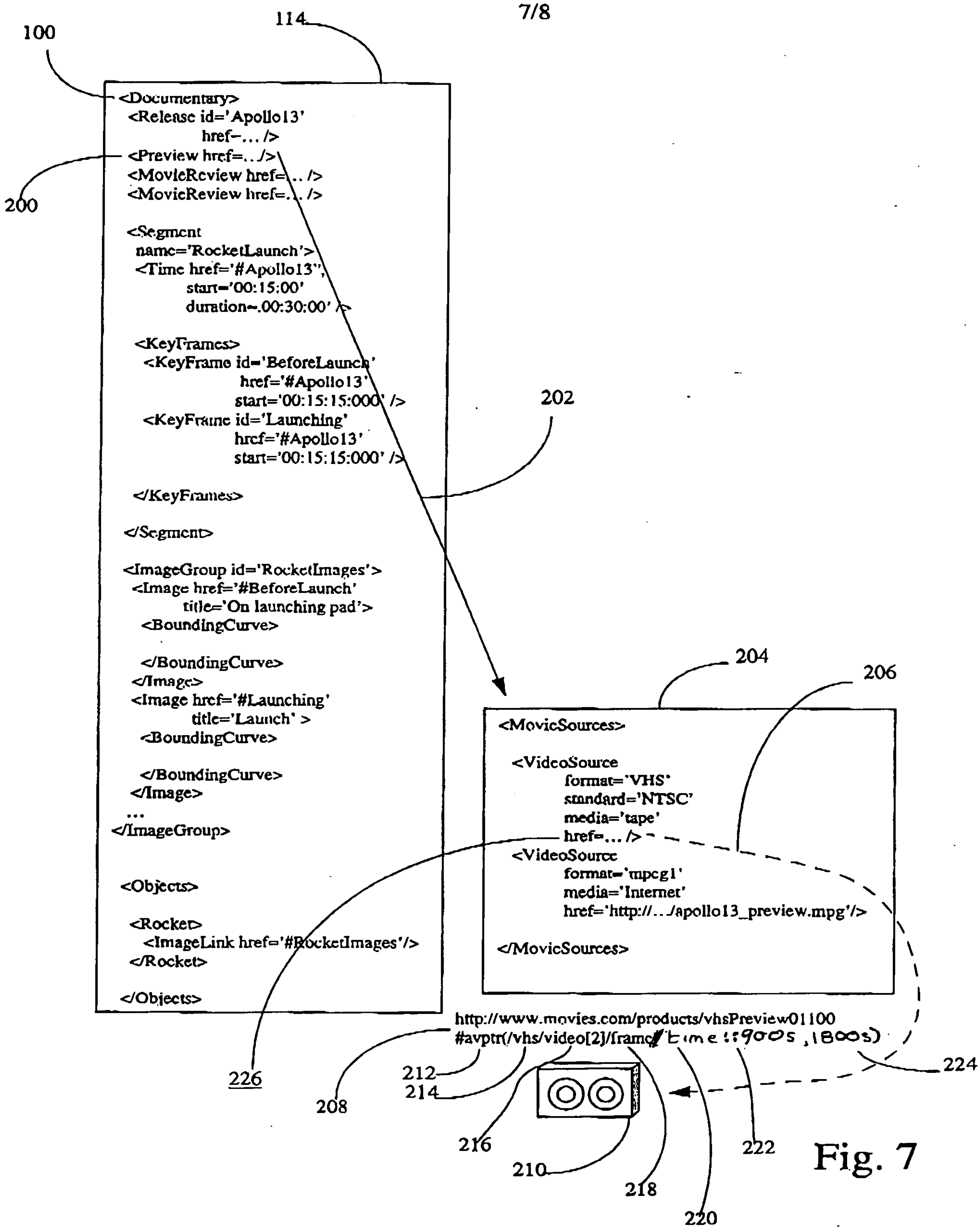

100
114
200
<Documentary>
<Release id='Apollo13'
href=... />
<Preview href=.../>
<MovieReview href=... />
<MovieReview href=... />
<Segment
name='RocketLaunch'>
<Time href='#Apollo13'
start='00:15:00'
duration=.00:30:00' />
<KeyFrames>
<KeyFrame id='BeforeLaunch'
href='#Apollo13'
start='00:15:15:000' />
<KeyFrame id='Launching'
href='#Apollo13'
start='00:15:15:000' />
</KeyFrames>
</Segment>
<ImageGroup id='RocketImages'>
<Image href='#BeforeLaunch'
title='On launching pad'>
<BoundingCurve>
</BoundingCurve>
</Image>
<Image href='#Launching'
title='Launch' >
<BoundingCurve>
</BoundingCurve>
</Image>
...
</ImageGroup>
<Objects>
<Rocket>
<ImageLink href='#RocketImages'/>
</Rocket>
</Objects>
202
204
206
<MovieSources>
<VideoSource
format='VHS'
standard='NTSC'
media='tape'
href=... />
<VideoSource
format='mpeg1'
media='Internet'
href='http://.../apollo13_preview.mpg'/>
</MovieSources>
http://www.movies.com/products/vhsPreview01100
#avptr(/vhs/video[2]/frame/time::900s,1800s)
226
208
212
214
216
210
218
220
222
224

Fig. 7

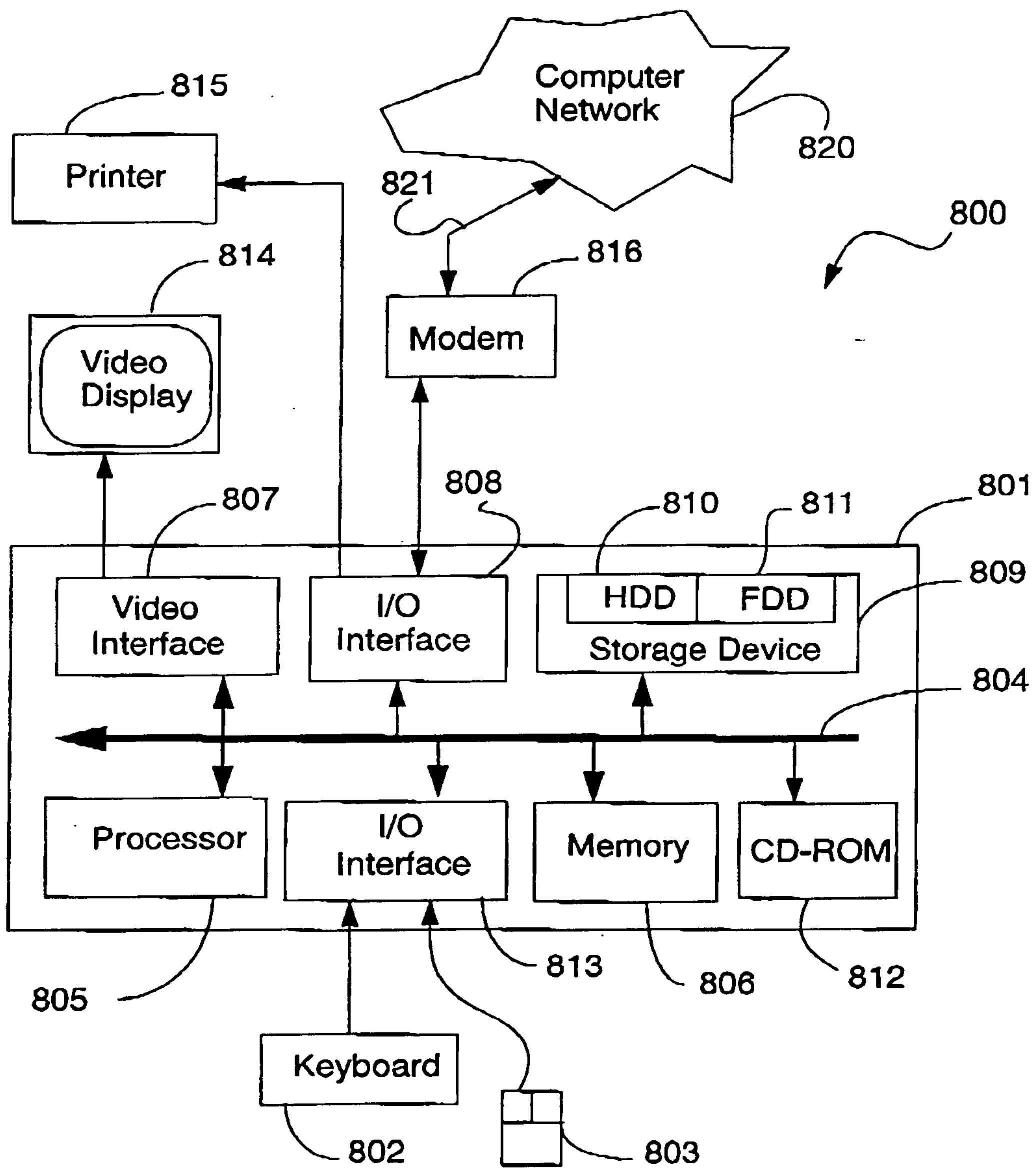

Computer
Network
820
815
Printer
821
800
816
814
Modem
Video
Display
808
807
810
811
801
809
HDD
FDD
Video
Interface
I/O
Interface
Storage Device
804
Processor
I/O
Interface
Memory
CD-ROM
805
813
806
812
Keyboard
802
803

Fig. 8